AGRICULTURAL ECONOMICS

Agricultural Economics 34 (2006) 273–280

Efficiency effects of agricultural economics research in the United States

David E. Schimmelpfennig*,[a], Christopher J. O'Donnell[b], George W. Norton[a,c]

[a]Economic Research Service, U.S. Department of Agriculture, 1800 M Street, NW, Room 4179, Washington DC 20036, USA
[b]Centre for Efficiency and Productivity Analysis (CEPA), School of Economics, University of Queensland, Australia
[c]Department of Agricultural and Applied Economics, Virginia Tech, Blacksburg, VA 24061, USA

Received 19 May 2004; received in revised form 14 January 2005; accepted 5 April 2005

Abstract

Allocations of research funds across programs are often made for efficiency reasons. Social science research is shown to have small, lagged but significant effects on U.S. agricultural efficiency when public agricultural R&D and extension are simultaneously taken into account. Farm management and marketing research variables are used to explain variations in estimates of allocative and technical efficiency using a Bayesian approach that incorporates stylized facts concerning lagged research impacts in a way that is less restrictive than popular polynomial distributed lags. Results are reported in terms of means and standard deviations of estimated probability distributions of parameters and long-run total multipliers. Extension is estimated to have a greater impact on both allocative and technical efficiency than either R&D or social science research.

JEL classification: Z00

Keywords: Social science research impacts; Bayesian estimation

1. Introduction

One of the primary areas of emphasis of the agricultural economics profession has been on assessing the benefits of production-oriented agricultural research and extension (ARE) (e.g., Bredahl and Peterson, 1976; Huffman and Evenson, 1992). Alston et al. (2000) and Evenson (2003) document the breadth and depth of this effort. Assessment studies have often provided the information needed to evaluate ARE for accountability purposes and to make resource allocation decisions across programs. More recently, efforts have been made to conceptualize and measure the benefits or impacts of social science research (SSR) in agriculture (Gardner, 2004; Lindner, 1987; Norton and Alwang, 2004). These studies suggest that the primary output of SSR is information. Thus, the problem of quantifying SSR impacts becomes a matter of valuing information. Economic surplus, decision theory, and econometric methods have all been considered for this purpose.

Economic surplus analysis (ESA) is particularly useful for valuing economic information from individual projects or well-defined programs (Alston et al., 1998). Combining ESA with decision theory may help in establishing causality between project-level SSR and eventual decisions by an economic agent (Gardner, 2004; Norton and Schuh, 1981; Schimmelpfennig and Norton, 2003). However, for accountability purposes, one often prefers to evaluate aggregate research programs rather than individual projects. For nonsocial science research, aggregate benefits of public and private ARE have been evaluated using econometric estimates of production, productivity, profit, and cost functions (e.g., Huffman and Evenson, 1993). A benefit of the econometric approach is that it provides a measure of the statistical reliability of the results. However, aggregate econometric assessment of SSR is difficult because of the diversity of the SSR programs affecting agriculture, the difficulty of separating out the effects of social science from those of other disciplines, and the fact that the users of social science research information are often one step removed from the beneficiaries.

For economic information that eventually affects producers, there have been a few econometric attempts to value the agricultural SSR input. For example, Norton (1987) assesses the impacts of farm management and marketing research and extension (MMRE) on improving technical (TE) and allocative (AE) efficiency in U.S. agriculture using a profit function. The rationale for this work is that TE measures the ability of the firm to minimize the inputs required to produce given outputs, and this is an information problem that SSR should be able to help solve. AE measures the inequality between the marginal rate of technical substitution for a pair of inputs and the ratio of their

*Corresponding author: Tel.: (202) 694-5507; fax: (202) 694-5775.
E-mail address: des@ers.usda.gov (D. E. Schimmelpfennig).

input prices—this also has strong information requirements and is potentially influenced by SSR.

In the past few years, several improvements have been made in procedures for estimating TE and AE, using both parametric and nonparametric approaches. The procedure used most widely in agricultural applications is the non-parametric data envelopment analysis (DEA) procedure developed by Charnes et al. (1994). This article uses DEA in a two-stage approach to assessing whether SSR related to agricultural marketing and management can explain variations in estimated efficiency across U.S. states. The first stage uses DEA to generate estimates of TE and AE. The second stage uses Bayesian methods to assess the impacts of MMRE on estimated TE and AE scores. Parametric approaches to obtaining TE and AE estimates, such as estimating profit or cost functions to obtain stochastic frontier estimates, require specification of a (possibly restrictive) functional form. This article uses the alternative DEA approach because it does not impose any prior restrictions on the underlying technology (Färe et al., 1985).

The next section uses the existing literature to motivate models that relate efficiency scores to variables including SSR. Having motivated the structure of the relationships, we then discuss estimation using sampling theory and Bayesian techniques. The following two sections then describe the data and the empirical results. In the concluding section, we relate our findings to previous results and discuss directions for future work.

2. Models and estimation methods

In order to relate efficiency scores to environmental variables such as SSR, the literature provides several options. Stochastic frontier analysts have accounted for environmental variables using both one-stage (Battese and Coelli, 1995) and two-stage approaches (Kalirajan, 1981; Pitt and Lee, 1981). The first stage in the two-stage approach involves estimating efficiency in a conventional stochastic frontier framework (i.e., without accounting for environmental variables), while the second stage involves regressing predicted efficiency scores on environmental variables. The problem with this approach is that the *predicted* efficiency scores can only be legitimately expressed as a function of the environmental variables if the latter are formally incorporated into the first stage (i.e., if we use a one-stage approach). Including environmental variables in the first stage of a stochastic frontier model makes the second stage unnecessary because the functional relationship between predicted efficiency scores and environmental variables is theoretically known. This is not the case for efficiency scores obtained using DEA—because the estimated frontier is nonparametric, the functional relationship between efficiency scores and environmental variables is unknown and no theoretical inconsistencies arise from using a two-stage approach.

The DEA approach can also be used to account for environmental variables in several ways (see Coelli et al., 1998). However, the two-stage procedure we use in this article is the most convenient method for accounting for dynamics in the relationship between DEA efficiency scores and SSR (in this article we confirm earlier results that point to significant lags between research spending and research impacts that are far beyond any normal agricultural input and output relationships). The two-stage procedure is also appealing in the context of previous work by Chavas and Cox (1992), Evenson et al. (1987), Evenson and Pray (1991), Huffman and Evenson (2001), McCunn and Huffman (2000), Schimmelpfennig and Thirtle (1999), and Thirtle and Bottomley (1989). These authors have established that U.S. agricultural R&D, extension (EX), and farmer education should be treated as determining variables in a second stage after the estimation of total factor productivity (TFP). This has been referred to as "two-stage decomposition."

One of the most significant lessons of previous two-stage decomposition analysis is that the models used for second-stage estimation should incorporate both lags of first-order variables and interaction terms. Due to the subtle and diffuse effects of MMRE on efficiency caused by the factors discussed above, it is difficult to obtain meaningful results without incorporating nonsample information (i.e., stylized facts gleaned from earlier studies) into the estimation process. The distributed lag models and estimation methods we use in this article have been developed in this context.

2.1. Distributed lag models

Previous research on ARE has yielded some stylized facts on research lag structures that we have attempted to incorporate. Agricultural R&D and EX alter the production environment facing producers. Public R&D is a public good with nonrival, nonexcludable characteristics, and therefore affects the production process as a whole, while EX has a broad mission to support "producers, families, communities, and other customers" (http://www.reeusda.gov/).[1] According to previous work (see references for two-stage decomposition) above, the effects of research are expected to influence AE and TE with lags of up to 30 years (peak effects at 15 years for the symmetric lag structure that is often assumed). EX should have shorter lags, skewed toward the first few years, and with smaller tails in its lag structure, but earlier empirical work has documented impacts of EX with up to 10-year lags. The topics of public R&D (probably more than EX) have changed over time but hopefully these changes mirror production changes (toward, e.g., more environment-friendly farming practices) with stable lag structures.

Previous authors have reasonably complained about the size of the task of testing all of these possible lags in research and EX impacts. We extend our own grief by adding an SSR variable and interaction terms. To deal with estimation problems we present a new estimation approach that has not been applied to ARE, and

[1] It might be argued that some aspects of agricultural extension influence the efficiency with which producers operate the production process, but our data are for the entire extension program.

which is neither as restrictive as polynomial distributed lags nor requires as many (implied) degrees of freedom as unrestricted estimation of individual lags.

For this new approach we developed several pieces of prior information based on previous research findings for the ARE lag weights:

1. The annual weights on research (R&D) are expected to increase in absolute value continuously up to at least lags five or six and then to decline in absolute value up to at least lag 15 (insufficient observations are available to allow testing longer lags, but this is expected to include peak effects).
2. EX is expected to have a shorter lag structure than research and to increase in absolute value up to lags 1 or 2 and then to decline in absolute value up to lag 10.
3. R&D and EX interaction terms with the same lag structure as EX are considered for their potential policy interest and total multiplier effects.
4. There is little evidence on expected lag weights for MMRE, but it seems likely that the short-run effect might be like EX with truncated longer-run effects—annual MMRE is expected to increase up to lags one or two and then decline up to lag five.

The result is the following two-equation model for the second-stage analysis:

$$\mathrm{TE}_{it} = \beta_1 + \sum_{j=0}^{p} \gamma_{1j}\mathrm{RD}_{i,t-j} + \sum_{j=0}^{q} \phi_{1j}\mathrm{EX}_{i,t-j} + \sum_{j=0}^{r} \alpha_{1j}\mathrm{MMRE}_{i,t-j} \sum_{j=0}^{s} \theta_{1j}\mathrm{RD}_{i,t-j}\mathrm{EX}_{i,t-j} + \varepsilon_{1it}, \quad (1)$$

$$\mathrm{AE}_{it} = \beta_2 + \sum_{j=0}^{p} \gamma_{2j}\mathrm{RD}_{i,t-j} + \sum_{j=0}^{q} \phi_{2j}\mathrm{EX}_{i,t-j} + \sum_{j=0}^{r} \alpha_{2j}\mathrm{MMRE}_{i,t-j} + \sum_{j=0}^{s} \theta_{2j}\mathrm{RD}_{i,t-j}\mathrm{EX}_{i,t-j} + \varepsilon_{2it}, \quad (2)$$

where the maximum lag lengths are $p = 15$, $q = 10$, $r = 5$, and $s = 10$. Consistent with the previous discussion, the lag coefficients satisfy the constraints:

$$\gamma_{n0} \le \gamma_{n1} \le \gamma_{n2} \le \gamma_{n3} \le \gamma_{n4} \le \gamma_{n5} \quad \text{and} \quad |\gamma_{n6}| \ge |\gamma_{n7}| \ge \cdots \ge |\gamma_{n15}|, \quad (3)$$

$$\phi_{n0} \le \phi_{n1} \text{ and } |\phi_{n2}| \ge |\phi_{n3}| \ge \cdots \ge |\phi_{n10}|, \quad (4)$$

$$\alpha_{n0} \le \alpha_{n1} \text{ and } |\alpha_{n2}| \ge |\alpha_{n3}| \ge \cdots \ge |\alpha_{n5}|, \quad \text{and} \quad (5)$$

$$\theta_{n0} \le \theta_{n1} \text{ and } |\theta_{n2}| \ge |\theta_{n3}| \ge \cdots \ge |\theta_{n10}|, \quad (6)$$

while the error terms satisfy $E\{\varepsilon_{jit}\} = 0$. For convenience, each of Eqs. (1) and (2) can be written more compactly as

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + e_t, \quad (7)$$

where $\boldsymbol{\beta}$ is 50×1, and $\mathbf{x}_t$ is $T \times 50$.[2] Unconstrained ordinary least squares (OLS) estimation of (7) is biased and inconsistent because the dependent variables have an upper limit of 1. Unbiased and consistent estimates can be obtained using a Tobit limited dependent-variable approach. In a Tobit framework, the observed variable y_t is determined by

$$y_t^* = \mathbf{x}_t'\boldsymbol{\beta} + e_t, \quad (8)$$

$$y_t = \begin{cases} 1 & \text{if } y_t^* > 1 \\ y_t^* & \text{otherwise,} \end{cases} \quad (9)$$

where e_t is iid normal with zero mean and variance $\boldsymbol{\sigma}^2$. Thus,

$$\Pr(y_t = 1) = \Pr(y_t^* \ge 1) = \Pr(\mathbf{x}_t'\boldsymbol{\beta} + e_t \ge 1) = \Phi\left(\frac{\mathbf{x}_t'\boldsymbol{\beta} - 1}{\boldsymbol{\sigma}}\right). \quad (10)$$

Moreover, the likelihood function is (see Maddala, 1983, p. 161):

$$p(\mathbf{y} \mid \boldsymbol{\theta}) = \prod_{t \le T_1} f_N\left(y_t|\mathbf{x}_t'\boldsymbol{\beta}, \boldsymbol{\sigma}^2\right) \times \prod_{t > T_1} \Phi\left(\frac{\mathbf{x}_t'\boldsymbol{\beta} - 1}{\boldsymbol{\sigma}}\right) \quad (11)$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma})$ and $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$. Estimating the parameters of (11) subject to the constraints (3)–(6) is not straightforward. Maximum likelihood estimation is possible if the absolute value signs are removed from the constraints. Alternatively, the absolute value signs can be retained and the model estimated in a Bayesian framework.

2.2. Maximum likelihood estimation

After dropping the absolute value signs from (3) to (6), we write the constraints in the form

$$\mathbf{D}\boldsymbol{\beta} \ge \mathbf{c}, \quad (12)$$

where $\mathbf{D}$ is a $K \times K$ nonsingular matrix and $\mathbf{c}$ is a $K \times 1$ vector with elements that are either 0 or $-\infty$. Using (12) the original model (7) can be rewritten as

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + e_t = \mathbf{x}_t'\mathbf{D}^{-1}\mathbf{D}\boldsymbol{\beta} + e_t = \mathbf{z}_t'\boldsymbol{\eta} + e_t, \quad (13)$$

where $\mathbf{z}_t{}' = \mathbf{x}_t{}'\mathbf{D}^{-1}$ and $\boldsymbol{\eta} = \mathbf{D}\boldsymbol{\beta} \ge \mathbf{c}$. Recall that, since the elements of $\mathbf{c}$ are either 0 or $-\infty$, the problem of estimating $\boldsymbol{\beta}$ in (7) subject to inequality constraints of the form (12) becomes one of estimating $\boldsymbol{\eta}$ in (13) subject to inequality constraints which ensure that particular elements of $\boldsymbol{\eta}$ are positive. In turn,

[2] Without pooling the data, the subscript t is used to index all observations (rather than using one subscript to index time periods and another subscript to index states). It is also convenient to order the observations so that the first T_1 observations on the dependent variable are uncensored, and the remaining $T - T_1$ observations are at their limit equal to 1.

constraining $\eta_k \geq 0$ can be accomplished by reparameterizing $\eta_k$as $\eta_k = \lambda_k^2$ and then estimating the model in terms of λ_k.

After making these alterations to (11) the likelihood function becomes

$$p(\mathbf{y} \mid \boldsymbol{\eta}, \boldsymbol{\sigma}^2) = \prod_{t \leq T_1} f_N\left(y_t | \mathbf{z}_t' \boldsymbol{\eta}, \boldsymbol{\sigma}^2\right) \times \prod_{t > T_1} \Phi\left(\frac{\mathbf{z}_t' \boldsymbol{\eta} - 1}{\boldsymbol{\sigma}}\right), \quad (14)$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$. Maximizing the likelihood function (14) is straightforward.

2.3. Bayesian estimation

Removing the absolute value signs from constraints (3) to (6) is computationally convenient for a sampling theory version of the model, but has the undesirable effect of not allowing negative coefficients to turn positive after peak effects (in addition to other undesirable side effects). To estimate the Tobit model subject to constraints on the relative magnitudes of the absolute values of the coefficients, it becomes necessary to use a Bayesian approach. Bayesian methods are becoming increasingly important in overcoming otherwise intractable sampling theory difficulties (Gao and Lahiri, 2000; O'Donnell et al., 1999).

To implement the Bayesian approach we adopt a noninformative joint prior (as in Chib, 1992, p. 89):

$$p(\boldsymbol{\theta}) \propto \frac{1}{\boldsymbol{\sigma}} \times I(\boldsymbol{\beta} \in S), \quad (15)$$

where $\propto$ denotes "proportional to" and $I(\cdot)$ is an indicator function, which takes the value 1 if the argument is true and 0 otherwise. S is the set of feasible values defined by constraints (3) to (6). The posterior density is then

$$\begin{aligned} p(\boldsymbol{\theta} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) = & \prod_{t \leq T_1} f_N\left(y_t | \mathbf{x}_t' \boldsymbol{\beta}, \boldsymbol{\sigma}^2\right) \\ & \times \prod_{t > T_1} \Phi\left(\frac{\mathbf{x}_t' \boldsymbol{\beta} - 1}{\boldsymbol{\sigma}}\right) \\ & \times \frac{1}{\sigma} \times I(\boldsymbol{\beta} \in S). \end{aligned} \quad (16)$$

Unfortunately, it is difficult to sample directly from this posterior using simple Markov Chain Monte Carlo (MCMC) algorithms. Instead, we follow Chib (1992) and Tanner and Wong (1987) and use a data augmentation algorithm that can be motivated by expressing the posterior as

$$\mathrm{p}(\boldsymbol{\theta} \mid \mathbf{y}) = \int p(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{z}) p(\mathbf{z} \mid \mathbf{y}) \, d\mathbf{z}, \quad (17)$$

where $p(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{z})$ is the posterior density of $\boldsymbol{\theta}$ given $\mathbf{y}$ and the vector of latent observations $\mathbf{z} = (y_{T_1+1}^*, y_{T_1+2,}^*, \ldots, y_T^*)'$. From Eq. (17) it can be seen that $p(\boldsymbol{\theta} \mid \mathbf{y})$ is the average of $p(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{z})$ over all possible values of $\mathbf{z}$. It follows that if we had a sample of M observations on $\mathbf{z}$, denoted $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(M)}$, we could approximate $p(\boldsymbol{\theta} \mid \mathbf{y})$ using

$$\hat{p}(\boldsymbol{\theta} \mid \mathbf{y}) = M^{-1} \sum_m p(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{z}^{(m)}). \quad (18)$$

This approximation could then be used to generate observations on $\boldsymbol{\theta}$, the parameter of interest. Thus, the first step is to obtain observations on $\mathbf{z}$. We do this by noting that the density $p(\mathbf{z} \mid \mathbf{y})$ can be expressed as

$$p(\mathbf{z} \mid \mathbf{y}) = \int p(\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta}. \quad (19)$$

Thus, if we had a sample of N observations on $\boldsymbol{\theta}$, denoted $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(N)}$, we could approximate $p(\mathbf{z} \mid \mathbf{y})$ as

$$\hat{p}(\mathbf{z} \mid \mathbf{y}) = N^{-1} \sum_n p(\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta}^{(n)}). \quad (20)$$

This approximation can be used to generate observations on $\mathbf{z}$. In summary, to generate observations on $\boldsymbol{\theta}$ and $\mathbf{z}$ we repeatedly follow these steps;

i) generate a sample of observations $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(M)}$ using an approximation to $p(\mathbf{z} \mid \mathbf{y})$,
ii) use $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(M)}$ to approximate $p(\boldsymbol{\theta} \mid \mathbf{y})$,
iii) generate a sample of observations $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(N)}$ using the current approximation to $p(\boldsymbol{\theta} \mid \mathbf{y})$,
iv) use $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(N)}$ to approximate $p(\mathbf{z} \mid \mathbf{y})$, and
v) generate a sample of observations $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(M)}$ using the current approximation to $p(\mathbf{z} \mid \mathbf{y})$.

Proceeding in this way, the sets of draws on $\boldsymbol{\theta}$ and $\mathbf{z}$ can be regarded as draws from the posterior densities $p(\boldsymbol{\theta} \mid \mathbf{y})$ and $p(\mathbf{z} \mid \mathbf{y})$. This result holds even when $M = N = 1$ (Tanner and Wong, 1987). In this article we make use of that result by setting $M = N = 1$, which allows us to draw observations on $\boldsymbol{\theta}$ and $\mathbf{z}$ by using

$$\begin{aligned} p(z_t | \mathbf{y}, \boldsymbol{\theta}) &= p(y_t^* | y_t^* \geq 1, \boldsymbol{\beta}, \boldsymbol{\sigma}^2) \\ &= f_N(y_t^* | \mathbf{x}_t' \boldsymbol{\beta}, \boldsymbol{\sigma}^2) / \boldsymbol{\Phi}\left(\frac{\mathbf{x}_t' \boldsymbol{\beta} - 1}{\boldsymbol{\sigma}}\right) \times I(y_t^* \geq 1), \end{aligned} \quad (21)$$

$$\begin{aligned} p(\boldsymbol{\beta} | \boldsymbol{\sigma}, \mathbf{z}, \mathbf{y}) \propto & f_N(\boldsymbol{\beta} | (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{y}^*, \boldsymbol{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}) \\ & \times I(\boldsymbol{\beta} \in S), \end{aligned} \quad (22)$$

$$\begin{aligned} p(\boldsymbol{\sigma} \mid \boldsymbol{\beta}, \mathbf{z}, \mathbf{y}) &\propto \mathbf{f}_{IG}(\boldsymbol{\sigma} | 2[(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})]^{-1}, T/2) \\ &\propto \tfrac{1}{\boldsymbol{\sigma}^{T+1}} \exp\left[-\tfrac{1}{2\boldsymbol{\sigma}^2}(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})\right], \end{aligned} \quad (23)$$

where $\mathbf{y}^* = (\mathrm{y}_1^*, \mathrm{y}_2^*, \ldots, \mathrm{y}_T^*) = (\mathrm{y}_1, \ldots, \mathrm{y}_{T_I}, \mathbf{z}')'$ denotes the full vector of latent variable observations and the inverted gamma notation, $f_{IG}(\cdot)$, is from Zellner (1971) (p. 371). The normal density in (21) follows from (10) and the assumption that e_t is normally distributed with mean 0 and variance $\boldsymbol{\sigma}^2$. The densities in (22) and (23) are standard results for multiple regression models with noninformative priors (Zellner, 1971, p. 66–67).

To generate from the truncated normal density in (21), we use Devroye (1986)[3]:

$$z_t = \mathbf{x}_t'\boldsymbol{\beta} + \boldsymbol{\sigma} \times \Phi^{-1}\left[\Phi\left(\frac{1-\mathbf{x}_t'\boldsymbol{\beta}}{\boldsymbol{\sigma}}\right) + U(0,1) \times \Phi\left(\frac{\mathbf{x}_t'\boldsymbol{\beta}-1}{\boldsymbol{\sigma}}\right)\right], \tag{24}$$

where $U(0, 1)$ denotes a standard uniform random variable. We can draw from (22) using a random-walk Metropolis–Hastings algorithm (e.g., Griffiths et al., 2000). Finally, to draw from (23) we simply note that if $\boldsymbol{\sigma}$ has an inverted gamma distribution with parameters a and b, then $\omega = 1/\boldsymbol{\sigma}^2$ has a gamma distribution with parameters a and b (Zellner, 1971, p. 371). Thus, we draw ω from $f_G(\omega \mid 2[(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})]^{-1}, T/2)$ and then obtain the draw $\boldsymbol{\sigma} = |\sqrt{\omega^{-1}}|$.

3. Data

The input- and output-related data used to estimate the model are state-level, transitive, multilateral Tornqvist indices for two outputs (livestock and crops) and four inputs (capital, labor, land, and materials). The data are available (accessed January, 2005) at http://usda.mannlib.cornell.edu after selecting "ERS" and "Agricultural Productivity in the US." No allowance is made for technical progress in calculating the DEA efficiency scores—this would require the estimation of a separate DEA frontier for each time period. To account for technical change, a time trend (T) representing Hicks-neutral technical change is included in the equation for TE and AE. The DEA estimates of AE and TE are for 48 contiguous states and 37 years (1960–1996) yielding a panel of 1,776 observations. With the lag structures discussed above, we end up with a sample of close to 1,000.

Several panels of "determining" or environmental variables are tested in the model. Data for total state-level agricultural R&D (RD) are from the Current Research Information System (CRIS) maintained by the Department of Agriculture.[4] The CRIS is an excellent source of disaggregated data on agricultural research expenditures and scientist numbers for the United States. The data for EX are real extension expenditures per farm by state obtained from the Cooperative State Research, Education, and Extension Service. State-level summary statistics for these variables and for MMRE are reported in Table 1. The MMRE variable itself is constructed from two categories in the CRIS database: Farm Business Management and Agricultural Marketing and Farm Supply. These two SSR variables are measured in full time equivalent (FTE) researchers, and are aggregated into one variable. The FTEs were converted to expenditures and subtracted from the RD variable to avoid double counting when that variable was constructed.[5]

Table 1
Descriptive statistics for agricultural R&D, extension, and social science research

Variable*	1960–1996			
	Mean	Standard deviation	Minimum	Maximum
RD	22.771	17.952	1.5612	111.44
EX	1.0380	2.4621	0.0109	23.517
MMRE	5.4106	3.9277	0.0**	25.200

*Notation refers to definitions in Eqs. (1) and (2).
**There was no measured spending on these categories of MMRE in some states in some years, for example, Rhode Island in 1975 and New Hampshire in 1986.

Several other variables might have immediate effects on the production process in each state. The coefficient of variation (CV) of crop and livestock prices was significant in Norton (1987). An output Herfindahl index (HH), a measure of the heterogeneity of the crop mix, is included in the current study because more agriculturally diverse states might be expected to find across-the-board efficiency more elusive. An infrastructure variable, the miles of roads built (HI), and annual state deviations from average rainfall (RN) are also included. Better roads would be expected to positively affect the potential efficiency of the production process, and more weather variability to negatively affect it. Farmer education is not considered because of allowances for educational attainment in the first-stage TFP data.

A factor that is unmeasured is research spillovers. Attempts to measure interstate spillovers of research knowledge, which might be expected to substitute for state RD, led to problems with multicollinearity and left too few degrees of freedom to make reliable inferences. Likewise, government farm support programs have not been considered because they showed little interstate variability. To economize on degrees of freedom and to attempt to capture some of these unmeasured effects, regional dummy variables were tested. However, these did not improve the estimated models.

4. Results

Bayesian parameter estimates and pseudo-t-ratios obtained from MCMC samples of size 500,000 are reported in Table 2.[6] The maximum likelihood estimates are not reported to save

[3] This method can be unreliable when the argument in square brackets is greater than 8 in absolute value. For such cases, an alternative method is provided by Geweke (1991).

[4] We thank Dennis Ungelsbee of USDA for making these data available to us.

[5] Unfortunately, an attempt to construct a management and marketing extension variable was thwarted by the numerous changes in the way extension FTEs and expenditures were categorized by the Cooperative State Research, Education, and Extension Service over time.

[6] Large samples were chosen because the Monte Carlo chains were slow-mixing. The term pseudo-t-ratio is used because, strictly speaking, t-ratios are not meaningful in a Bayesian context. Although it is a sampling theory concept, we will also use terms such as *statistical significance* in cases where point estimates are more than two estimated standard deviations from zero.

Table 2
Results of Bayesian estimation of impacts of research on technical and allocative efficiency—Eqs. (1) and (2)

Dependent variables	966 observations			
Regressors	Technical efficiency (TE)		Allocative efficiency (AE)	
	Coefficient	Pseudo-*t*-statistic	Coefficient	Pseudo-*t*-statistic
Constant	7.43E−01	3,905.66**	7.10E−01	241.23**
MMRE	−3.57E−05	−1.54	−6.36E−04	−1.80*
MMRE(−1)	1.11E−05	0.37	−1.87E−04	−1.01
MMRE(−2)	1.15E−04	1.86*	7.52E−04	2.65**
MMRE(−3)	7.45E−05	3.10**	3.09E−04	2.71**
MMRE(−4)	5.12E−05	3.25**	1.69E−04	2.25*
MMRE(−5)	−4.74E−06	−0.16	−1.06E−05	−0.17
RD	−2.47E−05	−2.60**	−3.21E−04	−3.27**
RD(−1)	−1.54E−05	−2.08*	−2.02E−04	−5.46**
RD(−2)	−8.05E−06	−1.29	−1.48E−04	−4.23**
RD(−3)	−1.76E−06	−0.35	−7.96E−05	−1.89*
RD(−4)	3.98E−06	0.82	−1.67E−05	−0.38
RD(−5)	1.40E−05	1.71*	7.63E−05	1.42
RD(−6)	2.39E−05	2.48**	3.60E−04	3.49**
RD(−7)	1.75E−05	2.53**	1.91E−04	3.09**
RD(−8)	1.34E−05	2.66**	1.51E−04	2.62**
RD(−9)	1.09E−05	2.54**	1.15E−04	3.11**
RD(−10)	8.88E−06	2.40**	9.10E−05	3.03**
RD(−11)	7.23E−06	2.29*	7.34E−05	2.76**
RD(−12)	5.84E−06	2.10*	5.98E−05	2.39**
RD(−13)	−4.44E−06	−1.66*	−4.57E−05	−2.06*
RD(−14)	−1.06E−06	−0.30	7.10E−06	0.20
RD(−15)	−1.85E−07	−0.09	−4.42E−06	−0.22
EX	−3.38E−04	−2.63**	−5.94E−03	−11.24**
EX(−1)	−2.19E−04	−1.90*	−4.79E−03	−5.16**
EX(−2)	2.31E−04	2.56**	4.13E−03	4.99**
EX(−3)	1.51E−04	2.62**	2.98E−03	4.07**
EX(−4)	1.06E−04	2.71**	2.09E−03	4.75**
EX(−5)	8.03E−05	2.60**	1.57E−03	4.25**
EX(−6)	6.28E−05	2.47**	1.23E−03	3.37**
EX(−7)	4.87E−05	2.20*	9.70E−04	2.65**
EX(−8)	3.61E−05	1.84*	−7.65E−04	−2.23*
EX(−9)	7.94E−06	0.28	−3.55E−04	−0.77
EX(−10)	1.62E−06	0.09	−6.43E−05	−0.23
RDEX	2.94E−05	12.67**	−2.10E−05	−0.40
RDEX(−1)	3.41E−05	14.15**	2.34E−05	0.68
RDEX(−2)	3.54E−05	15.51**	7.96E−05	2.32*
RDEX(−3)	3.33E−05	18.81**	5.81E−05	2.86**
RDEX(−4)	3.14E−05	24.74**	4.11E−05	3.22**
RDEX(−5)	3.01E−05	26.83**	2.93E−05	3.49**
RDEX(−6)	2.90E−05	28.32**	2.22E−05	3.20**
RDEX(−7)	2.78E−05	24.57**	1.69E−05	2.75**
RDEX(−8)	2.66E−05	20.36**	1.25E−05	2.20*
RDEX(−9)	2.52E−05	15.78**	−4.34E−06	−0.47
RDEX(−10)	2.24E−05	7.96**	−2.39E−06	−0.42
T	7.66E−04	135.97**	6.18E−04	14.66**
CV	−4.15E−02	−279.59**	5.30E−02	52.81**
HH	2.43E−01	692.66**	2.98E−01	50.47**
HI	2.39E−03	423.13**	−1.14E−03	−25.89**
RN	3.26E−04	53.26**	−6.34E−04	−4.68**

** and * denote significance at the 99 and 95% levels, respectively.

space (and also because they are obtained using constraints that are less than ideal). The results reported in Table 2 indicate that the social science variables (MMRE) have similar impacts on TE and AE. Farm Business Management, and Agricultural Marketing have significant and positive effects on both TE and AE in years 2, 3, and 4, with the size of the impact diminishing by half or more with each additional annual lag. MMRE has a significant and negative impact on AE in the current period, which could be caused by early adopters of social science innovations raising the bar for later adopters who come along 2–4 years later.

The effects of R&D and EX on TE and AE are also quite similar. R&D has positive impacts on TE in lags 5–12, and bigger positive impacts on AE in lags 6–12. R&D's impacts are negative on AE in the current period and lag one while they are negative on TE through lag three. The peak size effects of R&D on both TE and AE are in lag 6, declining each year thereafter until turning negative in lag 13.

The early effects of EX on both TE and AE are similarly negative in the current period and in lag one, showing a peak effect in lag two and declining and significant for six periods thereafter. Like those of MMRE, the early negative effects of ARE are probably caused by early adopting farmers operating on or near the efficiency frontier who make quick, thorough use of new technology and EX recommendations, to the relative detriment of most other farmers. It is unlikely that these are actual negative effects of ARE since their impacts are positive and significant for many years thereafter. The interaction terms indicate that R&D and EX are complements throughout, with effects on TE for 11 periods and effects on AE in periods 2–8. This complementarity of R&D and EX would tend to reinforce both the early negative and later positive individual effects of R&D and EX.

Price variability negatively influences TE but has a positive effect on AE, while the heterogeneity of the crop mix positively impacts both TE and AE. More miles of paved roads are associated with higher TE but lower AE, while higher deviations from average rainfall increase TE while reducing AE. Various reasons might be provided to explain these results, but the more revealing fact is that all of the supplementary variables are significant and three out of four of them have different signed effects on TE and AE. This probably indicates that the similarity of the results obtained for the research variables are not due to similarities in the construction of the DEA estimates, but are actual knowledge-based effects of research on efficiency.

4.1. Total multipliers

Most of the research variables have some significant negative lags so there is some interest in the size and sign of their total multipliers. To calculate a total multiplier for the effect of R&D on TE that accounts for the interaction with EX, we evaluate,

$$m_{11} = \frac{\partial \mathrm{TE}_{it}}{\partial \mathrm{RD}_{i,t}} + \frac{\partial \mathrm{TE}_{it}}{\partial \mathrm{RD}_{i,t-1}} + \cdots + \frac{\partial \mathrm{TE}_{it}}{\partial \mathrm{RD}_{i,t-p}} = \gamma_{10} + \gamma_{11} + \cdots + \gamma_{1,15} + \theta_{10}\mathrm{EX}_{it} + \theta_{11}\mathrm{EX}_{i,t-1} + \cdots + \theta_{1,10}\mathrm{EX}_{i,t-10}. \tag{25}$$

The total effect of social science research on TE can be measured more simply as,

$$m_{13} = \frac{\partial \mathrm{TE}_{it}}{\partial \mathrm{MMRE}_{i,t}} + \frac{\partial \mathrm{TE}_{it}}{\partial \mathrm{MMRE}_{i,t-1}} + \cdots + \frac{\partial \mathrm{TE}_{it}}{\partial \mathrm{MMRE}_{i,t-r}} = \alpha_{10} + \alpha_{12} + \cdots + \alpha_{15}, \quad (26)$$

since there are no estimated interaction effects between MMRE and ARE. Total multipliers are calculated similarly for each of the research variables on AE and all of the multipliers are evaluated by setting all current and lagged values of RD, EX, and MMRE equal to their sample means.

Bayesian estimates of the coefficients and pseudo-*t*-statistics for the long-run multipliers are reported in Table 3.[7] The total multipliers for MMRE show that farm management and marketing research have a positive and significant impact on both TE and AE. The mean effect of these social science research variables on TE is about one-half the size of their effect on AE. The overall effect of RD on TE, accounting for the complementarity between RD and EX previously identified in Table 2, is positive and roughly one-third smaller than the effect of RD on AE. These results indicate that agricultural R&D and social science research both improve producers' AE more than they increase the TE of the production system. This is consistent with the observation that crop mix decisions have become more complicated over the last 30+ years, as cropping alternatives have expanded. Research has focused on giving farmers more options to deal with changing market and environmental conditions and this research has tended to raise AE more than TE.

The overall effects of EX on TE and AE, including interaction effects with RD, are both positive and much larger in magnitude than the total effects of either RD or MMRE. Contrary to other research results, EX has a slightly bigger effect on TE than AE. This appears to indicate that EX is having fundamentally different impacts, working to increase the efficiency of the production system while other research has more important impacts on farmers' ability to effectively produce increasingly differentiated agricultural products. Substantial EX activities might focus on helping farmers benefit from farm programs, but these activities are unobservable in the present setting and would not be reflected in the AE estimates.

5. Conclusion

The value of agricultural marketing and management research in the United States appears to be positive, significant, and derived more from impacts on allocative than technical efficiency. The overall impact of SSR on allocative efficiency is almost twice as large as its impact on technical efficiency. A next step will be to use these results to generate a rate of return to public investments in this type of research, and to consider whether these results hold in other developed countries.

[7] Analogous equilibrium multipliers would simply be the sums of the lag coefficients for each variable (Greene, 1997, p. 784).

Table 3
Long-run total multipliers from Bayesian estimation of impacts of research on technical and allocative efficiency—Eqs. (1) and (2)

Dependent variables	966 observations			
Research variable	Technical efficiency (TE)		Allocative efficiency (AE)	
	Coefficient	Pseudo-*t*-statistic	Coefficient	Pseudo-*t*-statistic
MMRE ($m_{13,23}$)	2.11E−04	7.55**	3.97E−04	1.89*
RD ($m_{11,21}$)[a]	3.98E−04	79.11**	5.81E−04	8.47**
EX ($m_{12,22}$)	7.43E−03	210.77**	6.77E−03	32.93**

** and * denote significance at the 99 and 95% levels, respectively.
[a] For definitions of multipliers see Eqs. (25) and (26).

Public agricultural research and EX activities have positive overall effects on TE. R&D has a smaller positive effect that begins later, while EX has a strong positive impact from the second year and continuing into the higher lags. R&D depends to some extent on EX to disseminate new research results, but the size and longevity of EX's impacts on technical efficiency are substantial. R&D and EX both also have positive overall impacts on allocative efficiency. Even though the impact of R&D on AE is almost double its impact on TE, and EX's impact on AE is slightly smaller than on TE, EX's overall impact on AE is still several magnitudes greater than R&D's impact on AE. In contrast, Huffman and Evenson (2001) find that public crop research (unlagged stocks) have larger impacts on crop TFP than EX does (a familiar result), but we would expect EX and R&D to have different impacts on productivity than efficiency.

The availability of TE and AE estimates from the popular data envelopment analysis approach has allowed for a relatively straightforward assessment of the impacts of an important category of SSR, in addition to the impacts of agricultural R&D and EX. Studies of banking (Berger et al., 1993) and other sectors have found technical and allocative efficiency estimates to be somewhat dependent on the approach used to generate them (e.g., DEA, stochastic frontier cost or profit functions). Therefore, it may be useful to compare estimated impacts of SSR on TE and AE scores that have been generated with alternative techniques, to assess the robustness of the results.

Acknowledgments

Support for this research was provided in part through a Cooperative Agreement between the U.S. Department of Agriculture and Virginia Tech. The views expressed are not necessarily those of the U.S. Department of Agriculture.

References

Alston, J., Kang, C., Marra, M., Pardey, P., Wyatt, J., 2000. Meta-analysis of rates of return to agricultural R&D: Ex Pede Herculem. IFPRI Research Report 113, June, Washington DC.

Alston, J. M., Norton, G. W., Pardey, P. G., 1998. Science under Scarcity: Principles and Practice for Agricultural Research Evaluation and Priority Setting. CAB International, New York.

Battese, G. E., Coelli, T. J., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empir. Econ. 20, 325–332.

Berger, A. N., Hunter, W. C., Timme, S. G., 1993. The efficiency of financial institutions: a review and preview of research past, present and future. J Bank. Finance 17, 221–249.

Bredahl, M., Peterson, W. L., 1976. The productivity and allocation of research: U.S. Agricultural Experiment Stations. Am. J. Agric. Econ. 58(4) (Part I, Nov.), 684–692

Charnes, A., Cooper, W., Lewin, A., Seiford, L., 1994. Data Envelopment Analysis—Theory, Methodology and Applications. Kluwer, Boston.

Chavas, J. P., Cox, T. L., 1992. A Nonparametric analysis of the influence of research on agricultural productivity. Am. J. Agric. Econ. 74, 583–591.

Chib, S., 1992. Bayes inference in the Tobit censored regression model. J. Econometrics 5, 79–99.

Coelli, T., Rao, D. S. P., Battese, G. E., 1998. An Introduction to Efficiency and Productivity Analysis. Kluwer Academic, Boston, pp. 275.

Devroye, L., 1986. Non-Uniform Random Variate Generation. Springer-Verlag, New York.

Evenson, R. E., 2003. Private and public research and extension. In: Gardner, B. L., Rausser, G. C. (Eds.), Handbook of Agricultural Economics, Volume 1, Part 1. North-Holland/Elsevier Science Publishers, New York.

Evenson, R. E., Landau, D., Ballou, D., 1987. Agricultural productivity measurement for U.S. states. Evaluating Agricultural Research and Productivity, Miscellaneous Publication 52-1987. Agricultural Experiment Station, University of Minnesota, St. Paul, Minnesota.

Evenson, R. E., Pray, C. E., (Eds.) 1991. Research and Productivity in Asian Agriculture. Cornell University Press, Ithaca, NY.

Färe, R., Grosskopf, S., Lovell, C. A. K., 1985. The Measurement of Efficiency of Production. Kluwer Academic Publishers, Boston.

Gardner, B. L., 2004. Returns to policy-related social science research in agriculture. Chapter 9 In: Pardey, P. G., Smith, V. H. (Eds.), What's Economics Worth? Valuing Policy Research. Johns Hopkins University Press, Baltimore.

Gao, C., Lahiri, K., 2000. MCMC algorithms for two recent Bayesian limited information estimators. Econ. Lett. 66(2), 121–126.

Geweke, J., 1991. Efficient simulation from the multivariate normal and student t distributions subject to linear constraints. In: Keramidas, E. M., Kaufman, S. M. (Eds.), Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface. Interface Foundation of America, Seattle, pp. 571–578.

Greene, W. H., 1997. Econometric Analysis, 3rd ed. Prentice−Hall, Inc., Englewood Cliffs, New Jersey.

Griffiths, W. E., O'Donnell, C. J., Tan Cruz, A., 2000. Imposing regularity conditions on a system of cost and cost-share equations: a Bayesian approach. Aust. J. Agric. Econ. 44(1), 107–127.

Huffman, W. E., Evenson, R. E., 1992. Contributions of public and private science and technology to U.S. agricultural productivity. Am. J. Agric. Econ. 74, 751–756.

Huffman, W. E., Evenson, R. E., 1993. Science for Agriculture. Iowa State University Press, Ames, Iowa.

Huffman, W. E., Evenson, R. E., 2001. Structural and productivity change in U.S. agriculture, 1950-1982. Agric. Econ. 24, 127–147.

Kalirajan, K., 1981. An econometric analysis of yield variability in paddy production. Can. J. Agric. Econ. 29, 283–294.

Lindner, R., 1987. Toward a framework for evaluating agricultural economics research. Aust. J. Agric. Econ. Res. 31, 95–111.

Maddala, G. S., 1983. Limited Dependent and Qualitative Variables in Econometrics. CUP, Cambridge.

McCunn, A., Huffman, W. E., 2000. Convergence in U.S. productivity growth for agriculture: implications of interstate research spillovers for funding agricultural research. Am. J. Agric. Econ. 82(2), 370–388.

Norton, G. W., 1987. Evaluating social science research. In: Sundquist, W. B. (Ed.), Evaluating Agricultural Research and Productivity, Misc. Pub. No. 52-1987. University of Minnesota Agricultural Experiment Station, St. Paul, Minnesota.

Norton, G. W., Alwang, J., 2004. Measuring the benefits of policy-oriented social science research: evidence from two developing countries. Chapter 10 In: Pardey, P. G., Smith, V. H. (Eds.), What's Economics Worth? Valuing Policy Research. Johns Hopkins University Press, Baltimore.

Norton, G. W., Schuh, G. E., 1981. Evaluating returns to social science research: issues and possible methods. In: Norton, G. W., Fishel, W., Paulsen, A., Sundquist, W. B. (Eds.), Misc. Pub. No. 8-1981. University of Minnesota Agricultural Experiment Station, St. Paul, Minnesota.

O'Donnell, C. J., Shumway, C. R., Ball, V. E., 1999. Input demands and inefficiency in U.S. agriculture. Am. J. Agric. Econ. 81, 865–880.

Pitt, M. M., Lee, M-F., 1981. The measurement and sources of technical inefficiency in the Indonesian weaving industry. J. Dev. Econ. 9, 43–64.

Schimmelpfennig, D., Norton, G. W., 2003. What is the value of agricultural economics research? Am. J. Agric. Econ. 85(1), 81–94.

Schimmelpfennig, D., Thirtle, C., 1999. The internationalization of agricultural technology: patents, R&D spillovers and their effects on productivity in the European Union and United States. Contemp. Econ. Policy 17, 457–468.

Tanner, M. A., Wong, W-H., 1987. The calculation of posterior distributions by data augmentation. (with discussion) J. Am. Stat. Assoc. 82, 528–550.

Thirtle, C., Bottomley, P., 1989. The rate of return to public sector agricultural R&D in the UK, 1965-80. Appl. Econ. 21, 1063–1086.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. John Wiley, New York.